DTM-Based Filtrations

Hirokazu Anai
Fujitsu Laboratories, AI Lab, Kawasaki, Japan
anai@jp.fujitsu.com

Frédéric Chazal
Datashape, Inria Paris-Saclay, France
frederic.chazal@inria.fr

Marc Glisse
Datashape, Inria Paris-Saclay, France
marc.glisse@inria.fr

Yuichi Ike
Fujitsu Laboratories, AI Lab, Kawasaki, Japan
ike.yuichi@jp.fujitsu.com

Hiroya Inakoshi
Fujitsu Laboratories, AI Lab, Kawasaki, Japan
hiroya.inakoshi@jp.fujitsu.com

Raphaël Tinarrage
Datashape, Inria Paris-Saclay, France
raphael.tinarrage@inria.fr

Yuhei Umeda
Fujitsu Laboratories, AI Lab, Kawasaki, Japan
umeda.yuhei@jp.fujitsu.com

Abstract

Despite strong stability properties, the persistent homology of filtrations classically used in Topological Data Analysis, such as, e.g. the Čech or Vietoris-Rips filtrations, are very sensitive to the presence of outliers in the data from which they are computed. In this paper, we introduce and study a new family of filtrations, the DTM-filtrations, built on top of point clouds in the Euclidean space which are more robust to noise and outliers. The approach adopted in this work relies on the notion of distance-to-measure functions and extends some previous work on the approximation of such functions.

2012 ACM Subject Classification Theory of computation → Computational geometry

Keywords and phrases Topological Data Analysis, Persistent homology

Digital Object Identifier 10.4230/LIPIcs.SoCG.2019.58

Related Version The complete version of the paper, including proofs and additional comments, can be found at `https://arxiv.org/abs/1811.04757`.

Funding This work was partially supported by a collaborative research agreement between Inria and Fujitsu, and the Advanced Grant of the European Research Council GUDHI (Geometric Understanding in Higher Dimensions).

1 Introduction

The inference of relevant topological properties of data represented as point clouds in Euclidean spaces is a central challenge in Topological Data Analysis (TDA).

Given a (finite) set of points X in $\mathbb{R}^d$, persistent homology provides a now classical and powerful tool to construct persistence diagrams whose points can be interpreted as

35th International Symposium on Computational Geometry (SoCG 2019).
Editors: Gill Barequet and Yusu Wang; Article No. 58; pp. 58:1–58:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

homological features of X at different scales. These persistence diagrams are obtained from *filtrations*, i.e. nested families of subspaces or simplicial complexes, built on top of X. Among the many filtrations available to the user, unions of growing balls $\cup_{x\in X}\overline{B}(x,t)$ (sublevel sets of distance functions), $t \in \mathbb{R}^+$, and their nerves, the Čech complex filtration, or its usually easier to compute variation, the Vietoris-Rips filtration, are widely used. The main theoretical advantage of these filtrations is that they have been shown to produce persistence diagrams that are stable with respect to perturbations of X in the Hausdorff metric [6].

Unfortunately, the Hausdorff distance turns out to be very sensitive to noise and outliers, preventing the direct use of distance functions and classical Čech or Vietoris-Rips filtrations to infer relevant topological properties from real noisy data. Several attempts have been made in the recent years to overcome this issue. Among them, the filtration defined by the sublevel sets of the distance-to-measure (DTM) function introduced in [4], and some of its variants [10], have been proven to provide relevant information about the geometric structure underlying the data. Unfortunately, from a practical perspective, the exact computation of the sublevel sets filtration of the DTM, that boils down to the computation of a k-th order Voronoï diagram, and its persistent homology turn out to be far too expensive in most cases. To address this problem, [8] introduces a variant of the DTM function, the witnessed k-distance, whose persistence is easier to compute and proves that the witnessed k-distance approximates the DTM persistence up to a fixed additive constant. In [3, 2], a weighted version of the Vietoris-Rips complex filtration is introduced to approximate the persistence of the DTM function, and several stability and approximation results, comparable to the ones of [8], are established. Another kind of weighted Vietoris-Rips complex is presented in [1].

Contributions. In this paper, we introduce and study a new family of filtrations based on the notion of DTM. Our contributions are the following:

- Given a set $X \subset \mathbb{R}^d$, a weight function f defined on X and $p \in [1, +\infty]$, we introduce the weighted Čech and Rips filtrations that extend the notion of sublevel set filtration of power distances of [3]. Using classical results, we show that these filtrations are stable with respect to perturbations of X in the Hausdorff metric and perturbations of f with respect to the sup norm (Propositions 3 and 4).
- For a general function f, the stability results of the weighted Čech and Rips filtrations are not suited to deal with noisy data or data containing outliers. We consider the case where f is the empirical DTM-function associated to the input point cloud. In this case, we show an outliers-robust stability result: given two point clouds $X, Y \subseteq \mathbb{R}^d$, the closeness between the persistence diagrams of the resulting filtrations relies on the existence of a subset of X which is both close to X and Y in the Wasserstein metric (Theorems 15 and 20).

Practical motivations. Even though this aspect is not considered in this paper, it is interesting to mention that the DTM filtration was first experimented in the setting of an industrial research project whose goal was to address an anomaly detection problem from inertial sensor data in bridge and building monitoring [9]. In this problem, the input data comes as time series measuring the acceleration of devices attached to the monitored bridge/building. Using sliding windows and time-delay embedding, these times series are converted into a series of fixed size point clouds in $\mathbb{R}^d$. Filtrations are then built on top of these point clouds and their persistence is computed, giving rise to a time-dependent sequence of persistence diagrams that are then used to detect anomalies or specific features occurring along the time [11, 13]. In this practical setting it turned out that the DTM filtrations reveal to be not only more resilient to noise but also able to better highlight topological features in the data than the standard Vietoris-Rips filtrations, as illustrated on a basic synthetic example on Figure 1. One of the goals of the present work is to provide theoretical foundations to these promising experimental results by studying the stability properties of the DTM filtrations.

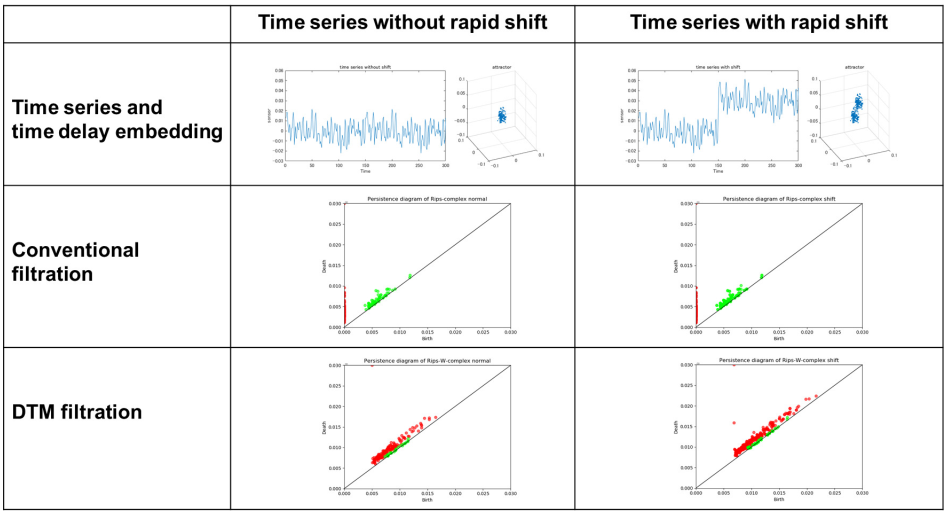


Figure 1 A synthetic example comparing Vietoris-Rips filtration to DTM filtration. The first row represents two time series with very different behavior and their embedding into $\mathbb{R}^3$ (here a series $(x_1, x_2, \dots, x_n)$ is converted in the 3D point cloud $\{(x_1, x_2, x_3), (x_2, x_3, x_4), \dots, (x_{n-2}, x_{n-1}, x_n)\}$). The second row shows the persistence diagrams of the Vietoris-Rips filtration built on top of the two point clouds (red and green points represent respectively the 0-dimensional 1-dimensional diagrams); one observes that the diagrams do not clearly 'detect' the different behavior of the time series. The third row shows the persistence diagrams of the DTM filtration built on top of the two point clouds; a red point clearly appears away from the diagonal in the second diagram that highlights the rapid shift occurring in the second time series.

Organisation of the paper. Preliminary definitions, notations, and basic notions on filtrations and persistence modules are recalled in Section 2. The weighted Čech and Vietoris-Rips filtrations are introduced in Section 3, where their stability properties are established. The DTM-filtrations are introduced in Section 4. Their main stability properties are established in Theorems 15 and 20, and their relation with the sublevel set filtration of the DTM-functions is established in Proposition 16.

The various illustrations and experiments of this paper have been computed with the GUDHI library on Python [14].

For the complete version of this paper, including proofs and additional comments, see the online version at `https://arxiv.org/abs/1811.04757`.

2 Filtrations and interleaving distance

In the sequel, we consider interleavings of filtrations, interleavings of persistence modules and their associated pseudo-distances. Their definitions, restricted to the setting of the paper, are briefly recalled in this section.

Let $T = \mathbb{R}^+$ and $E = \mathbb{R}^d$ endowed with the standard Euclidean norm.

Filtrations of sets and simplicial complexes. A family of subsets $(V^t)_{t\in T}$ of $E = \mathbb{R}^d$ is a *filtration* if it is non-decreasing for the inclusion, i.e. for any $s, t \in T$, if $s \leq t$ then $V^s \subseteq V^t$. Given $\epsilon \geq 0$, two filtrations $(V^t)_{t\in T}$ and $(W^t)_{t\in T}$ of E are *ϵ-interleaved* if, for every $t \in T$, $V^t \subseteq W^{t+\epsilon}$ and $W^t \subseteq V^{t+\epsilon}$. The interleaving pseudo-distance between $(V^t)_{t\in T}$ and $(W^t)_{t\in T}$ is defined as the infimum of such ϵ:

$$d_i((V^t)_{t\in T}, (W^t)_{t\in T}) = \inf\{\epsilon : (V^t) \text{ and } (W^t) \text{ are } \epsilon\text{-interleaved}\}.$$

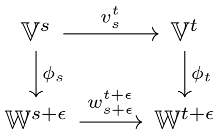
Filtrations of simplicial complexes and their interleaving distance are similarly defined: given a set X and an abstract simplex S with vertex set X, a *filtration of S* is a non-decreasing family $(S^t)_{t\in T}$ of subcomplexes of S. The interleaving pseudo-distance between two filtrations $(S_1^t)_{t\in T}$ and $(S_2^t)_{t\in T}$ of S is the infimum of the $\epsilon \geq 0$ such that they are ϵ-interleaved, i.e. for any $t \in T$, $S_1^t \subseteq S_2^{t+\epsilon}$ and $S_2^t \subseteq S_1^{t+\epsilon}$.

Notice that the interleaving distance is only a pseudo-distance, as two distinct filtrations may have zero interleaving distance.

Persistence modules. Let k be a field. A *persistence module* $\mathbb{V}$ over $T = \mathbb{R}^+$ is a pair $\mathbb{V} = ((\mathbb{V}^t)_{t\in T}, (v_s^t)_{s\leq t\in T})$ where $(\mathbb{V}^t)_{t\in T}$ is a family of k-vector spaces, and $(v_s^t : \mathbb{V}^s \to \mathbb{V}^t)_{s\leq t\in T}$ a family of linear maps such that:

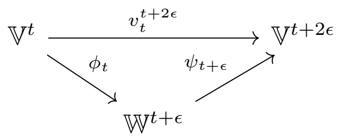
- for every $t \in T$, $v_t^t : V^t \to V^t$ is the identity map,
- for every $r, s, t \in T$ such that $r \leq s \leq t$, $v_s^t \circ v_r^s = v_r^t$.

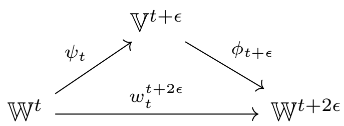
Given $\epsilon \geq 0$, an *ϵ-morphism* between two persistence modules $\mathbb{V}$ and $\mathbb{W}$ is a family of linear maps $(\phi_t : \mathbb{V}^t \to \mathbb{W}^{t+\epsilon})_{t\in T}$ such that the following diagrams commute for every $s \leq t \in T$:



If $\epsilon = 0$ and each ϕ_t is an isomorphism, the family $(\phi_t)_{t\in T}$ is said to be an *isomorphism* of persistence modules.

An *ϵ-interleaving* between two persistence modules $\mathbb{V}$ and $\mathbb{W}$ is a pair of ϵ-morphisms $(\phi_t : \mathbb{V}^t \to \mathbb{W}^{t+\epsilon})_{t\in T}$ and $(\psi_t : \mathbb{W}^t \to \mathbb{V}^{t+\epsilon})_{t\in T}$ such that the following diagrams commute for every $t \in T$:

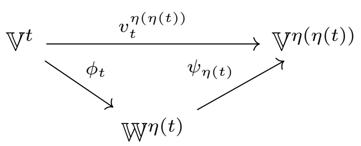


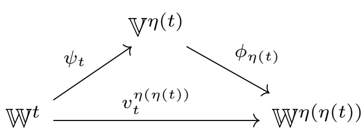


The interleaving pseudo-distance between $\mathbb{V}$ and $\mathbb{W}$ is defined as

$$d_i(\mathbb{V}, \mathbb{W}) = \inf\{\epsilon \geq 0, \mathbb{V} \text{ and } \mathbb{W} \text{ are } \epsilon\text{-interleaved}\}.$$

In some cases, the proximity between persistence modules is expressed with a function. Let $\eta : T \to T$ be a non-decreasing function such that for any $t \in T$, $\eta(t) \geq t$. A η-interleaving between two persistence modules $\mathbb{V}$ and $\mathbb{W}$ is a pair of families of linear maps $(\phi_t : \mathbb{V}^t \to \mathbb{W}^{\eta(t)})_{t\in T}$ and $(\psi_t : \mathbb{W}^t \to \mathbb{V}^{\eta(t)})_{t\in T}$ such that the following diagrams commute for every $t \in T$:





When η is $t \mapsto t + c$ for some $c \geq 0$, it is called an additive c-interleaving and corresponds with the previous definition. When η is $t \mapsto ct$ for some $c \geq 1$, it is called a multiplicative c-interleaving.

A persistent module $\mathbb{V}$ is said to be *q-tame* if for every $s, t \in T$ such that $s < t$, the map v_s^t is of finite rank. The q-tameness of a persistence module ensures that we can define a notion of persistence diagram – see [5]. Moreover, given two q-tame persistence modules $\mathbb{V}, \mathbb{W}$ with persistence diagrams $D(\mathbb{V}), D(\mathbb{W})$, the so-called isometry theorem states that $d_b(D(\mathbb{V}), D(\mathbb{W})) = d_i(\mathbb{V}, \mathbb{W})$ ([5, Theorem 4.11]) where $d_b(\cdot, \cdot)$ denotes the bottleneck distance between diagrams.

Relation between filtrations and persistence modules. Applying the homology functor to a filtration gives rise to a persistence module where the linear maps between homology groups are induced by the inclusion maps between sets (or simplicial complexes). As a consequence, if two filtrations are ϵ-interleaved then their associated homology persistence modules are also ϵ-interleaved, the interleaving homomorphisms being induced by the interleaving inclusion maps. Moreover, if the modules are q-tame, then the bottleneck distance between their persistence diagrams is upperbounded by ϵ.

The filtrations considered in this paper are obtained as union of growing balls. Their associated persistence module is the same as the persistence module of a filtered simplicial complex via the persistent nerve lemma ([7], Lemma 3.4). Indeed, consider a filtration $(V^t)_{t\in T}$ of E and assume that there exists a family of points $(x_i)_{i\in I} \in E^I$ and a family of non-decreasing functions $r_i : T \to \mathbb{R}^+ \cup \{-\infty\}$, $i \in I$, such that, for every $t \in T$, V^t is equal to the union of closed balls $\bigcup_I \overline{B}(x_i, r_i(t))$, with the convention $\overline{B}(x_i, -\infty) = \emptyset$. For every $t \in T$, let $\mathcal{V}^t$ denote the cover $\{\overline{B}(x_i, r_i(t)), i \in I\}$ of V^t, and S^t be its nerve. Let $\mathbb{V}$ be the persistence module associated with the filtration $(V^t)_{t\in T}$, and $\mathbb{V}_{\mathcal{N}}$ the one associated with the simplicial filtration $(S^t)_{t\in T}$. Then $\mathbb{V}$ and $\mathbb{V}_{\mathcal{N}}$ are isomorphic persistence modules. In particular, if $\mathbb{V}$ is q-tame, $\mathbb{V}$ and $\mathbb{V}_{\mathcal{N}}$ have the same persistence diagrams.

3 Weighted Čech filtrations

In order to define the DTM-filtrations, we go through an intermediate and more general construction, namely the weighted Čech filtrations. It generalizes the usual notion of Čech filtration of a subset of $\mathbb{R}^d$, and shares comparable regularity properties.

3.1 Definition

In the sequel of the paper, the Euclidean space $E = \mathbb{R}^d$, the index set $T = \mathbb{R}^+$ and a real number $p \geq 1$ are fixed. Consider $X \subseteq E$ and $f : X \to \mathbb{R}^+$. For every $x \in X$ and $t \in T$, we define

$$r_x(t) = \begin{cases} -\infty & \text{if } t < f(x), \\ \left(t^p - f(x)^p\right)^{\frac{1}{p}} & \text{otherwise.} \end{cases}$$

We denote by $\overline{B}_f(x, t) = \overline{B}(x, r_x(t))$ the closed Euclidean ball of center x and radius $r_x(t)$. By convention, a Euclidean ball of radius $-\infty$ is the empty set. For $p = \infty$, we also define

$$r_x(t) = \begin{cases} -\infty & \text{if } t < f(x), \\ t & \text{otherwise,} \end{cases}$$

and the balls $\overline{B}_f(x, t) = \overline{B}(x, r_x(t))$. Some of these radius functions are represented in Figure 2.

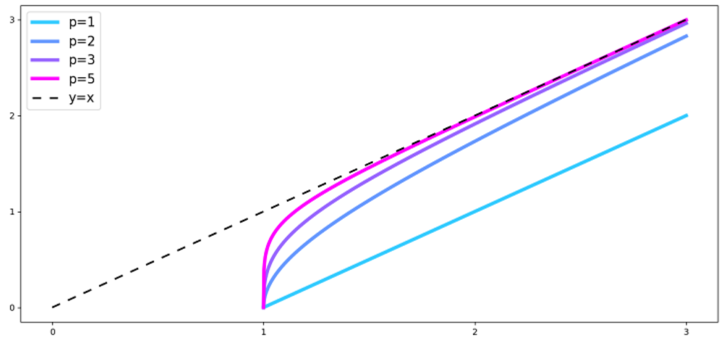


Figure 2 Graph of $t \mapsto r_x(t)$ for $f(x) = 1$ and several values of p.

▶ **Definition 1.** *Let $X \subseteq E$ and $f : X \to \mathbb{R}^+$. For every $t \in T$, we define the following set:*

$$V^t[X, f] = \bigcup_{x \in X} \overline{B}_f(x, t).$$

The family $V[X, f] = (V^t[X, f])_{t \geq 0}$ is a filtration of E. It is called the weighted Čech filtration with parameters (X, f, p). *We denote by $\mathbb{V}[X, f]$ its persistent (singular) homology module.*

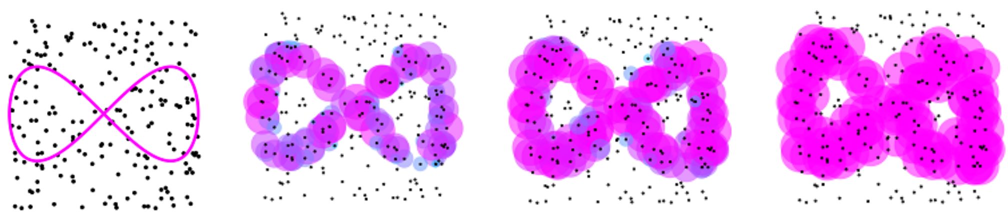
Note that $V[X, f]$ and $\mathbb{V}[X, f]$ depend on fixed parameter p, that is not made explicit in the notation.

Introduce $\mathcal{V}^t[X, f] = \{\overline{B}_f(x, t)\}_{x \in X}$. It is a cover of $V^t[X, f]$ by closed Euclidean balls. Let $\mathcal{N}(\mathcal{V}^t[X, f])$ be the nerve of the cover $\mathcal{V}^t[X, f]$. It is a simplicial complex over the vertex set X. The family $\mathcal{N}(\mathcal{V}[X, f]) = (\mathcal{N}(\mathcal{V}^t[X, f]))_{t \geq 0}$ is a filtered simplicial complex. We denote by $\mathbb{V}_{\mathcal{N}}[X, f]$ its persistent (simplicial) homology module. As a consequence of the persistent nerve theorem [7, Lemma 3.4], $\mathbb{V}[X, f]$ and $\mathbb{V}_{\mathcal{N}}[X, f]$ are isomorphic persistent modules.

When $f = 0$, $V[X, f]$ does not depend on $p \geq 1$, and it is the filtration of E by the sublevel sets of the distance function to X. In the sequel, we denote it by $V[X, 0]$. The corresponding filtered simplicial complex, $\mathcal{N}(\mathcal{V}[X, 0])$, is known as the usual Čech complex of X.

When $p = 2$, the filtration value of $y \in E$, i.e. the infimum of the t such that $y \in V^t[X, f]$, is called the power distance of y associated to the weighted set (X, f) in [3, Definition 4.1]. The filtration $V[X, f]$ is called the weighted Čech filtration ([3, Definition 5.1]).

Example. Consider the point cloud X drawn on the left of Figure 3 (black). It is a 200-sample of the uniform distribution on $[-1, 1]^2 \subseteq \mathbb{R}^2$. We choose f to be the distance function to the lemniscate of Bernoulli (magenta). Let $t = 0.2$. Figure 3 represents the sets $V^t[X, f]$ for several values of p. The balls are colored differently according to their radius.



X $\quad$ $p = 1$ $\quad$ $p = 2$ $\quad$ $p = \infty$

Figure 3 The set X and the sets $V^t[X, f]$ for $t = 0.2$ and several values of p.

The following proposition states the regularity of the persistent module $\mathbb{V}[X, f]$.

▶ **Proposition 2.** *If $X \subseteq E$ is finite and f is any function, then $\mathbb{V}[X, f]$ is a pointwise finite-dimensional persistence module.*

More generally, if X is a bounded subset of E and f is any function, then $\mathbb{V}[X, f]$ is q-tame.

3.2 Stability

We still consider a subset $X \subseteq E$ and a function $f : X \to \mathbb{R}^+$. Using the fact that two ϵ-interleaved filtrations induce ϵ-interleaved persistence modules, the stability results for the filtration $V[X, f]$ of this subsection immediately translate as stability results for the persistence module $\mathbb{V}[X, f]$.

The following proposition relates the stability of the filtration $V[X, f]$ with respect to f.

▶ **Proposition 3.** *Let $g : X \to \mathbb{R}^+$ be a function such that $\sup_{x\in X} |f(x) - g(x)| \leq \epsilon$. Then the filtrations $V[X, f]$ and $V[X, g]$ are ϵ-interleaved.*

The following proposition states the stability of $V[X, f]$ with respect to X. It generalizes [3, Proposition 4.3] (case $p = 2$).

▶ **Proposition 4.** *Let $Y \subseteq E$ and suppose that $f : X \cup Y \to \mathbb{R}^+$ is c-Lipschitz, $c \geq 0$. Suppose that X and Y are compact and that the Hausdorff distance $d_H(X, Y) \leq \epsilon$. Then the filtrations $V[X, f]$ and $V[Y, f]$ are k-interleaved with $k = \epsilon(1 + c^p)^{\frac{1}{p}}$.*

One can show that the bounds in Proposition 3 and 4 are tight.

When considering data with outliers, the observed set X may be very distant from the underlying signal Y in Hausdorff distance. Therefore, the tight bound in Proposition 4 may be unsatisfactory. Moreover, a usual choice of f would be d_X, the distance function to X. But the bound in Proposition 3 then becomes $\|d_X - d_Y\|_\infty = d_H(X, Y)$. We address this issue in Section 4 by considering an outliers-robust function f, the so-called distance-to-measure function (DTM).

3.3 Weighted Vietoris-Rips filtrations

Rather than computing the persistence of the Čech filtration of a point cloud $X \subseteq E$, one sometimes consider the corresponding Vietoris-Rips filtration, which is usually easier to compute.

If G is a graph with vertex set X, its corresponding clique complex is the simplicial complex over X consisting of the sets of vertices of cliques of G. If S is a simplicial complex, its corresponding flag complex is the clique complex of its 1-skeleton.

Recall that $\mathcal{N}(\mathcal{V}^t[X, f])$ denotes the nerve of $\mathcal{V}^t[X, f]$, where $\mathcal{V}^t[X, f]$ is the cover $\{\overline{B}_f(x, t)\}_{x\in X}$ of $V^t[X, f]$.

▶ **Definition 5.** *We denote by $\text{Rips}(\mathcal{V}^t[X, f])$ the flag complex of $\mathcal{N}(\mathcal{V}^t[X, f])$, and by $\text{Rips}(\mathcal{V}[X, f])$ the corresponding filtered simplicial complex. It is called the weighted Rips complex with parameters (X, f, p).*

The following proposition states that the filtered simplicial complexes $\mathcal{N}(\mathcal{V}[X, f])$ and $\text{Rips}(\mathcal{V}[X, f])$ are 2-interleaved multiplicatively, generalizing the classical case of the Čech and Vietoris-Rips filtrations (case $f = 0$).

▶ **Proposition 6.** *For every $t \geq 0$,*

$$\mathcal{N}(\mathcal{V}^t[X,f]) \subseteq \mathrm{Rips}(\mathcal{V}^t[X,f]) \subseteq \mathcal{N}(\mathcal{V}^{2t}[X,f])$$

Using Theorem 3.1 of [1], the multiplicative interleaving $\mathrm{Rips}(\mathcal{V}^t[X,f]) \subseteq \mathcal{N}(\mathcal{V}^{2t}[X,f])$ can be improved to $\mathrm{Rips}(\mathcal{V}^t[X,f]) \subseteq \mathcal{N}(\mathcal{V}^{ct}[X,f])$, where $c = \sqrt{\frac{2d}{d+1}}$ and d is the dimension of the ambient space $E = \mathbb{R}^d$.

Note that weighted Rips filtration shares the same stability properties as the weighted Čech filtration. Indeed, the proofs of Proposition 3 and 4 immediately extend to this case.

In order to compute the flag complex $\mathrm{Rips}(\mathcal{V}^t[X,f])$, it is enough to know the filtration values of its 0- and 1-simplices. The following proposition describes these values.

▶ **Proposition 7.** *Let $p < +\infty$. The filtration value of a vertex $x \in X$ is given by $t_X(\{x\}) = f(x)$.*
The filtration value of an edge $\{x,y\} \subseteq E$ is given by

$$t_X(\{x,y\}) = \begin{cases} \max\{f(x), f(y)\} & \text{if } \|x-y\| \leq |f(x)^p - f(y)^p|^{\frac{1}{p}}, \\ t & \text{otherwise,} \end{cases}$$

where t is the only positive root of

$$\|x-y\| = (t^p - f(x)^p)^{\frac{1}{p}} + (t^p - f(y)^p)^{\frac{1}{p}} \tag{1}$$

When $\|x-y\| \geq |f(x)^p - f(y)^p|^{\frac{1}{p}}$, the positive root of Equation (1) does not always admit a closed form. We give some particular cases for which it can be computed.

- For $p = 1$, the root is $t_X(\{x,y\}) = \frac{f(x)+f(y)+\|x-y\|}{2}$,
- for $p = 2$, it is $t_X(\{x,y\}) = \frac{\sqrt{\left((f(x)+f(y))^2+\|x-y\|^2\right)\left((f(x)-f(y))^2+\|x-y\|^2\right)}}{2\|x-y\|}$,
- for $p = \infty$, the condition reads $\|x-y\| \geq \max\{f(x), f(y)\}$, and the root is $t_X(\{x,y\}) = \frac{\|x-y\|}{2}$. In either case, $t_X(\{x,y\}) = \max\{f(x), f(y), \frac{\|x-y\|}{2}\}$.

We close this subsection by discussing the influence of p on the weighted Čech and Rips filtrations. Let $D_0(\mathcal{N}(\mathcal{V}[X,f,p]))$ be the persistence diagram of the 0th-homology of $\mathcal{N}(\mathcal{V}[X,f,p])$. We say that a point (b,d) of $D_0(\mathcal{V}[X,f,p])$ is non-trivial if $b \neq d$. Let $D_0(\mathrm{Rips}(\mathcal{V}[X,f,p]))$ be the persistence diagram of the 0th-homology of $\mathrm{Rips}(\mathcal{V}[X,f,p])$. Note that $D_0(\mathcal{N}(\mathcal{V}[X,f,p])) = D_0(\mathrm{Rips}(\mathcal{V}[X,f,p]))$ since the corresponding filtrations share the same 1-skeleton.

▶ **Proposition 8.** *The number of non-trivial points in $D_0(\mathrm{Rips}(\mathcal{V}[X,f,p]))$ is non-increasing with respect to $p \in [1,+\infty)$. The same holds for $D_0(\mathcal{N}(\mathcal{V}[X,f,p]))$.*

Figure 7 in Subsection 4.4 illustrates the previous proposition in the case of the DTM-filtrations. Greater values of p lead to sparser 0th-homology diagrams.

Now, consider $k > 0$, and let $D_k(\mathcal{N}(\mathcal{V}[X,f,p]))$ be the persistence diagram of the kth-homology of $\mathcal{N}(\mathcal{V}[X,f,p])$. In this case, one can easily build examples showing that the number of non-trivial points of $D_k(\mathcal{N}(\mathcal{V}[X,f,p]))$ does not have to be non-increasing with respect to p. The same holds for $D_k(\mathrm{Rips}(\mathcal{V}[X,f,p]))$.

4 DTM-filtrations

The results of previous section suggest that in order to construct a weighted Čech filtration $V[X, f]$ that is robust to outliers, it is necessary to choose a function f that depends on X and that is itself robust to outliers. The so-called distance-to-measure function (DTM) satisfies such properties, motivating the introduction of the DTM-filtrations in this section.

4.1 The distance to measure (DTM)

Let μ be a probability measure over $E = \mathbb{R}^d$, and $m \in [0, 1)$ a parameter. For every $x \in \mathbb{R}^d$, let $\delta_{\mu,m}$ be the function defined on E by $\delta_{\mu,m}(x) = \inf\{r \geq 0, \mu(\overline{B}(x, r)) > m\}$.

▶ **Definition 9.** *Let $m \in [0, 1[$. The DTM μ of parameter m is the function:*

$$\begin{array}{rccc} d_{\mu,m}: & E & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & \sqrt{\frac{1}{m}\int_0^m \delta_{\mu,t}^2(x)dt} \end{array}$$

When m is fixed – which is the case in the following subsections – and when there is no risk of confusion, we write d_μ instead of $d_{\mu,m}$.

Notice that when $m = 0$, $d_{\mu,m}$ is the distance function to $\text{supp}(\mu)$, the support of μ.

▶ **Proposition 10** ([4], Corollary 3.7)**.** *For every probability measure μ and $m \in [0, 1)$, $d_{\mu,m}$ is 1-Lipschitz.*

A fundamental property of the DTM is its stability with respect to the probability measure μ in the Wasserstein metric. Recall that given two probability measures μ and ν over E, a transport plan between μ and ν is a probability measure π over $E \times E$ whose marginals are μ and ν. The *Wasserstein distance with quadratic cost between μ and ν* is defined as $W_2(\mu, \nu) = \left(\inf_\pi \int_{E\times E} \|x - y\|^2 d\pi(x, y)\right)^{\frac{1}{2}}$, where the infimum is taken over all the transport plans π. When $\mu = \mu_X$ and $\nu = \mu_Y$ are the empirical measures of the finite point clouds X and Y, i.e the normalized sums of the Dirac measures on the points of X and Y respectively, we write $W_2(X, Y)$ instead of $W_2(\mu_X, \mu_Y)$.

▶ **Proposition 11** ([4], Theorem 3.5)**.** *Let μ, ν be two probability measures, and $m \in (0, 1)$. Then*

$$\|d_{\mu,m} - d_{\nu,m}\|_\infty \leq m^{-\frac{1}{2}} W_2(\mu, \nu).$$

Notice that for every $x \in E$, $d_\mu(x)$ is not lower than the distance from x to $\text{supp}(\mu)$, the support of μ. This remark, along with the propositions 10 and 11, are the only properties of the DTM that will be used to prove the results in the sequel of the paper.

In practice, the DTM can be computed. If X is a finite subset of E of cardinal n, we denote by μ_X its empirical measure. Assume that $m = \frac{k_0}{n}$, with k_0 an integer. In this case, $d_{\mu_X,m}$ reformulates as follows: for every $x \in E$,

$$d_{\mu_X,m}^2(x) = \frac{1}{k_0}\sum_{k=1}^{k_0} \|x - p_k(x)\|^2,$$

where $p_1(x), ..., p_{k_0}(x)$ are a choice of k_0-nearest neighbors of x in X.

4.2 DTM-filtrations

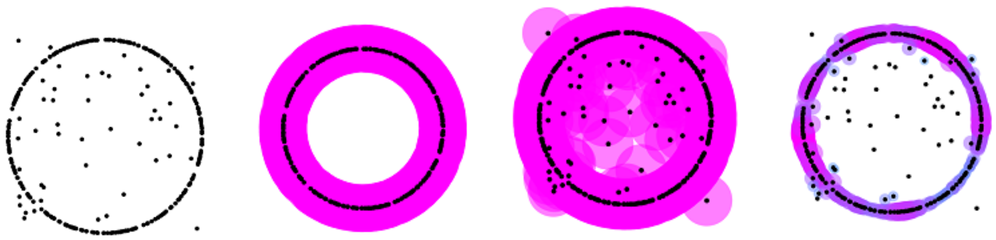
In the following, the two parameters $p \in [1, +\infty]$ and $m \in (0, 1)$ are fixed.

▶ **Definition 12.** *Let $X \subseteq E$ be a finite point cloud, μ_X the empirical measure of X, and d_{μ_X} the corresponding DTM of parameter m. The weighted Čech filtration $V[X, d_{\mu_X}]$, as defined in Definition 1, is called the* DTM-filtration associated with the parameters (X, m, p). *It is denoted by $W[X]$. The corresponding persistence module is denoted by $\mathbb{W}[X]$.*

Let $\mathcal{W}^t[X] = \mathcal{V}^t[X, d_{\mu_X}]$ denote the cover of $W^t[X]$ as defined in section 3, and let $\mathcal{N}(\mathcal{W}^t[X])$ be its nerve. The family $\mathcal{N}(\mathcal{W}[X])) = (\mathcal{N}(\mathcal{W}^t[X]))_{t \geq 0}$ is a filtered simplicial complex, and its persistent (simplicial) homology module is denoted by $\mathbb{W}_{\mathcal{N}}[X]$. By the persistent nerve lemma, the persistence modules $\mathbb{W}[X]$ and $\mathbb{W}_{\mathcal{N}}[X]$ are isomorphic.

As in Definition 5, $\mathrm{Rips}(\mathcal{W}^t[X])$ denotes the flag complex of $\mathcal{N}(\mathcal{W}^t[X])$, and $\mathrm{Rips}(\mathcal{W}[X])$ the corresponding filtered simplicial complex.

Example. Consider the point cloud X drawn on the left of Figure 4. It is the union of $\widetilde{X}$ and Γ, where $\widetilde{X}$ is a 50-sample of the uniform distribution on $[-1, 1]^2 \subseteq \mathbb{R}^2$, and Γ is a 300-sample of the uniform distribution on the unit circle. We consider the weighted Čech filtrations $V[\Gamma, 0]$ and $V[X, 0]$, and the DTM-filtration $W[X]$, for $p = 1$ and $m = 0.1$. They are represented in Figure 4 for the value $t = 0.3$.

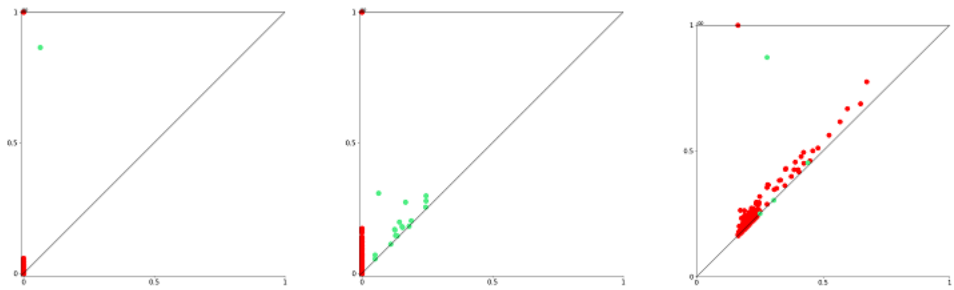


X $\quad V^t[\Gamma, 0]$ $\quad V^t[X, 0]$ $\quad W^t[X]$

Figure 4 The set X and the sets $V^t[\Gamma, 0]$, $V^t[X, 0]$ and $W^t[X]$ for $p = 1$, $m = 0.1$ and $t = 0.3$.

Because of the outliers $\widetilde{X}$, the value of t from which the sets $V^t[X, 0]$ are contractible is small. On the other hand, we observe that the set $W^t[X]$ does not suffer too much from the presence of outliers.

We plot in Figure 5 the persistence diagrams of the persistence modules associated to $\mathrm{Rips}(\mathcal{V}[\Gamma, 0])$, $\mathrm{Rips}(\mathcal{V}[X, 0])$ and $\mathrm{Rips}(\mathcal{W}[X])$ ($p = 1$, $m = 0.1$).



$D(\mathrm{Rips}(\mathcal{V}[\Gamma, 0]))$ $\quad D(\mathrm{Rips}(\mathcal{V}[X, 0]))$ $\quad D(\mathrm{Rips}(\mathcal{W}[X]))$

Figure 5 Persistence diagrams of some simplicial filtrations. Points in red (resp. green) represent the persistent homology in dimension 0 (resp. 1).

Observe that the diagrams $D(\mathrm{Rips}(\mathcal{V}[\Gamma, 0]))$ and $D(\mathrm{Rips}(\mathcal{W}[X]))$ appear to be close to each other, while $D(\mathrm{Rips}(\mathcal{V}[X, 0]))$ does not.

Applying the results of Section 3, we immediately obtain the following proposition.

▶ **Proposition 13.** *Consider two measures μ, ν on E with compact supports X and Y. Then*

$$d_i(V[X, d_\mu], V[Y, d_\nu]) \leq m^{-\frac{1}{2}} W_2(\mu, \nu) + 2^{\frac{1}{p}} d_H(X, Y).$$

In particular, if X and Y are finite subsets of E, using $\mu = \mu_X$ and $\nu = \nu_Y$, we obtain

$$d_i(W[X], W[Y]) \leq m^{-\frac{1}{2}} W_2(X, Y) + 2^{\frac{1}{p}} d_H(X, Y).$$

Note that this stability result is worse than the stability of the usual Čech filtrations, which only involves the Hausdorff distance: $d_i(V[X, 0], V[Y, 0]) \leq d_H(X, Y)$. However, the term $W_2(X, Y)$ is inevitable.

In the case where the Hausdorff distance $d_H(X, Y)$ is small, it would be more robust to consider these usual Čech filtrations. However, in the case where is it large, the usual Čech filtrations may be very distant. On the other hand, the DTM-filtrations may still be close, as we discuss in the next subsection.

4.3 Stability when $p = 1$

We first consider the case $p = 1$, for which the proofs are simpler and results are stronger. We fix $m \in (0, 1)$. If μ is a probability measure on E with compact support $\mathrm{supp}(\mu)$, we define

$$c(\mu, m) = \sup_{\mathrm{supp}(\mu)} (d_{\mu,m}).$$

If $\mu = \mu_\Gamma$ is the empirical measure of a finite set $\Gamma \subseteq E$, we denote it $c(\Gamma, m)$.

▶ **Proposition 14.** *Let μ be a probability measure on E with compact support Γ. Let d_μ be the corresponding DTM. Consider a set $X \subseteq E$ such that $\Gamma \subseteq X$. The weighted Čech filtrations $V[\Gamma, d_\mu]$ and $V[X, d_\mu]$ are $c(\mu, m)$-interleaved.*

Moreover, if $Y \subseteq E$ is another set such that $\Gamma \subseteq Y$, $V[X, d_\mu]$ and $V[Y, d_\mu]$ are $c(\mu, m)$-interleaved.

In particular, if Γ is a finite set and $\mu = \mu_\Gamma$ its empirical measure, $W[\Gamma]$ and $V[X, d_{\mu_\Gamma}]$ are $c(\Gamma, m)$-interleaved.

▶ **Theorem 15.** *Consider two measures μ, ν on E with supports X and Y. Let μ', ν' be two measures with compact supports Γ and Ω such that $\Gamma \subseteq X$ and $\Omega \subseteq Y$. We have*

$$d_i(V[X, d_\mu], V[Y, d_\nu]) \leq m^{-\frac{1}{2}} W_2(\mu, \mu') + m^{-\frac{1}{2}} W_2(\mu', \nu') + m^{-\frac{1}{2}} W_2(\nu', \nu) + c(\mu', m) + c(\nu', m).$$

In particular, if X and Y are finite, we have

$$d_i(W[X], W[Y]) \leq m^{-\frac{1}{2}} W_2(X, \Gamma) + m^{-\frac{1}{2}} W_2(\Gamma, \Omega) + m^{-\frac{1}{2}} W_2(\Omega, Y) + c(\Gamma, m) + c(\Omega, m).$$

Moreover, with $\Omega = Y$, we obtain

$$d_i(W[X], W[\Omega]) \leq m^{-\frac{1}{2}} W_2(X, \Gamma) + m^{-\frac{1}{2}} W_2(\Gamma, \Omega) + c(\Gamma, m) + c(\Omega, m).$$

The last inequality of Theorem 15 can be seen as an approximation result. Indeed, suppose that Ω is some underlying set of interest, and X is a sample of it with, possibly, noise or outliers. If one can find a subset Γ of X such that X and Γ are close to each other – in the Wasserstein metric – and such that Γ and Ω are also close, then the filtrations $W[X]$

and $W[\Omega]$ are close. Their closeness depends on the constants $c(\Gamma, m)$ and $c(\Omega, m)$. More generally, if X is finite and μ' is a measure with compact support $\Omega \subset X$ not necessarily finite, note that the first inequality gives

$$d_i(W[X], V[\Omega, d_{\mu'}]) \leq m^{-\frac{1}{2}} W_2(X, \Gamma) + m^{-\frac{1}{2}} W_2(\mu_\Gamma, \mu') + c(\Gamma, m) + c(\mu', m).$$

For any probability measure μ of support $\Gamma \subseteq E$, the constant $c(\mu, m)$ might be seen as a bias term, expressing the behaviour of the DTM over Γ. It relates to the concentration of μ on its support. Recall that a measure μ with support Γ is said to be (a, b)-standard, with $a, b \geq 0$, if for all $x \in \Gamma$ and $r \geq 0$, $\mu(\overline{B}(x, r)) \geq \min\{ar^b, 1\}$. For example, the Hausdorff measure associated to a compact b-dimensional submanifold of E is (a, b)-standard for some $a > 0$. In this case, a simple computation shows that there exists a constant C, depending only on a and b, such that for all $x \in \Gamma$, $d_{\mu,m}(x) \leq Cm^{\frac{1}{b}}$. Therefore, $c(\mu, m) \leq Cm^{\frac{1}{b}}$.

Regarding the second inequality of Theorem 15, suppose for the sake of simplicity that one can choose $\Gamma = \Omega$. The bound of Theorem 15 then reads

$$d_i(W[X], W[Y]) \leq m^{-\frac{1}{2}} W_2(X, \Gamma) + m^{-\frac{1}{2}} W_2(\Gamma, Y) + 2c(\Gamma, m).$$

Therefore, the DTM-filtrations $W[X]$ and $W[Y]$ are close to each other if μ_X and μ_Y are both close to a common measure μ_Γ. This would be the case if X and Y are noisy samples of Γ. This bound, expressed in terms of Wasserstein distance rather than Hausdorff distance, shows the robustness of the DTM-filtration to outliers.

Notice that, in practice, for finite data sets X, Y and for given Γ and Ω, the constants $c(\Gamma, m)$ and $c(\Omega, m)$ can be explicitly computed, as it amounts to evaluating the DTM on Γ and Ω. This remark holds for the bounds of Theorem 15.

Example. Consider the set $X = \widetilde{X} \cup \Gamma$ as defined in the example page 10. Figure 6 displays the sets $W^t[X]$, $V^t[X, d_{\mu_\Gamma}]$ and $W^t[\Gamma]$ for the values $p = 1$, $m = 0.1$ and $t = 0.4$ and the persistence diagrams of the corresponding weighted Rips filtrations, illustrating the stability properties of Proposition 14 and Theorem 15.

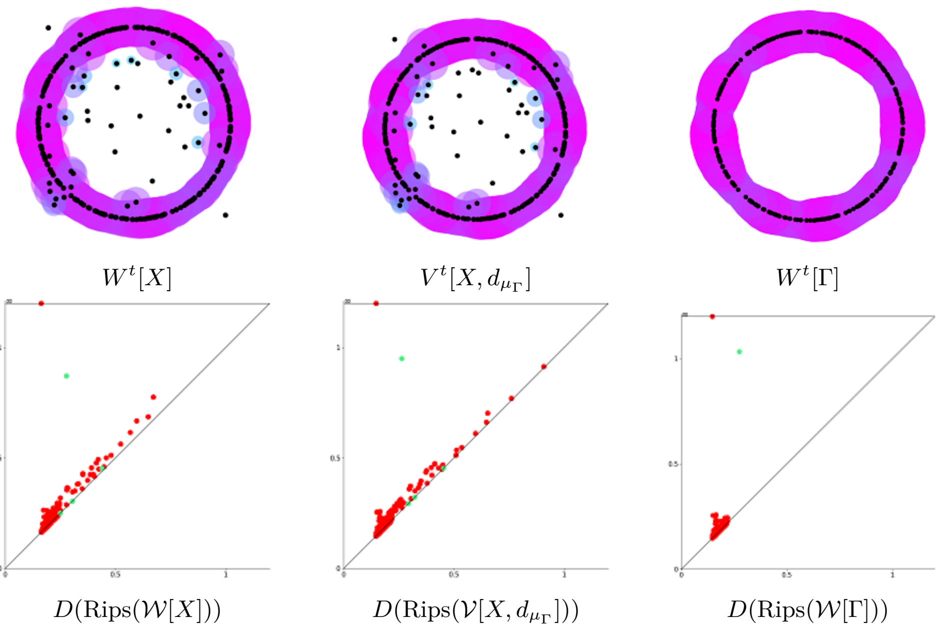


Figure 6 Filtrations for $t = 0.4$, and their corresponding persistence diagrams.

The following proposition relates the DTM-filtration to the filtration of E by the sublevels sets of the DTM.

▶ **Proposition 16.** *Let μ be a probability measure on E with compact support K. Let $m \in [0,1)$ and denote by V the sublevel sets filtration of d_μ. Consider a finite set $X \subseteq E$. Then*

$$d_i(V, W[X]) \leq m^{-\frac{1}{2}} W_2(\mu, \mu_X) + 2\epsilon + c(\mu, m),$$

with $\epsilon = d_H(K \cup X, X)$.

As a consequence, one can use the DTM-filtration to approximate the persistent homology of the sublevel sets filtration of the DTM, which is expensive to compute in practice.

We close this subsection by noting that a natural strengthening of Theorem 15 does not hold: let $m \in (0,1)$ and $E = \mathbb{R}^n$ with $n \geq 1$. There is no constant C such that, for every probability measure μ, ν on E with supports X and Y, we have:

$$d_i(V[X, d_{\mu,m}], V[Y, d_{\nu,m}]) \leq CW_2(\mu, \nu).$$

The same goes for the weaker statement

$$d_i(\mathbb{V}[X, d_{\mu,m}], \mathbb{V}[Y, d_{\nu,m}]) \leq CW_2(\mu, \nu).$$

4.4 Stability when $p > 1$

Now assume that $p > 1$, $m \in (0,1)$ being still fixed. In this case, stability properties turn out to be more difficult to establish. For small values of t, Lemma 18 gives a tight non-additive interleaving between the filtrations. However, for large values of t, the filtrations are poorly interleaved. To overcome this issue we consider stability at the homological level, i.e. between the persistence modules associated to the DTM filtrations.

If μ is a probability measure on E with compact support Γ, we define

$$c(\mu, m, p) = \sup_\Gamma(d_{\mu,m}) + \kappa(p) t_\mu(\Gamma),$$

where $\kappa(p) = 1 - \frac{1}{p}$, and $t_\mu(\Gamma)$ is the filtration value of the simplex Γ in $\mathcal{N}(\mathcal{V}[\Gamma, d_\mu])$, the (simplicial) weighted Čech filtration. Equivalently, $t_\mu(\Gamma)$ is the value t from which all the balls $\overline{B}_{d_\mu}(\gamma, t)$, $\gamma \in \Gamma$, share a common point.
If $\mu = \mu_\Gamma$ is the empirical measure of a finite set $\Gamma \subseteq E$, we denote it $c(\Gamma, m, p)$.

Note that we have the inequality $\frac{1}{2}\mathrm{diam}(\Gamma) \leq t_\mu(\Gamma) \leq 2\mathrm{diam}(\Gamma)$.

▶ **Proposition 17.** *Let μ be a measure on E with compact support Γ, and d_μ be the corresponding DTM of parameter m. Consider a set $X \subseteq E$ such that $\Gamma \subseteq X$. The persistence modules $\mathbb{V}[\Gamma, d_\mu]$ and $\mathbb{V}[X, d_\mu]$ are $c(\mu, m, p)$-interleaved.*

Moreover, if $Y \subseteq E$ is another set such that $\Gamma \subseteq Y$, $\mathbb{V}[X, d_\mu]$ and $\mathbb{V}[Y, d_\mu]$ are $c(\mu, m, p)$-interleaved.

In particular, if Γ is a finite set and $\mu = \mu_\Gamma$ its empirical measure, $\mathbb{W}[\Gamma]$ and $\mathbb{V}[X, d_{\mu_\Gamma}]$ are $c(\Gamma, m, p)$-interleaved.

The proof involves the two following ingredients. The first lemma gives a (non-additive) interleaving between the filtrations $W[\Gamma]$ and $V[X, d_{\mu_\Gamma}]$, relevant for low values of t, while the second proposition gives a result for large values of t.

▶ **Lemma 18.** *Let μ, Γ and X be as defined in Proposition 17. Let $\phi : t \mapsto 2^{1-\frac{1}{p}}t + \sup_\Gamma d_\mu$. Then for every $t \geq 0$,*

$$V^t[\Gamma, d_\mu] \subseteq V^t[X, d_\mu] \subseteq V^{\phi(t)}[\Gamma, d_\mu].$$

▶ **Proposition 19.** *Let μ, Γ and X be as defined in Proposition 17. Consider the map $v_*^t : \mathbb{V}^t[X, d_\mu] \to \mathbb{V}^{t+c}[X, d_\mu]$ induced in homology by the inclusion $v^t : V^t[X, d_\mu] \to V^{t+c}[X, d_\mu]$. If $t \geq t_\mu(\Gamma)$, then v^t is trivial.*

▶ **Theorem 20.** *Consider two measures μ, ν on E with supports X and Y. Let μ', ν' be two measures with compact supports Γ and Ω such that $\Gamma \subseteq X$ and $\Omega \subseteq Y$. We have*

$$d_i(\mathbb{V}[X, d_\mu], \mathbb{V}[Y, d_\nu]) \leq m^{-\frac{1}{2}} W_2(\mu, \mu') + m^{-\frac{1}{2}} W_2(\mu', \nu') + m^{-\frac{1}{2}} W_2(\nu', \nu) + c(\mu', m, p) + c(\nu', m, p).$$

In particular, if X and Y are finite, we have

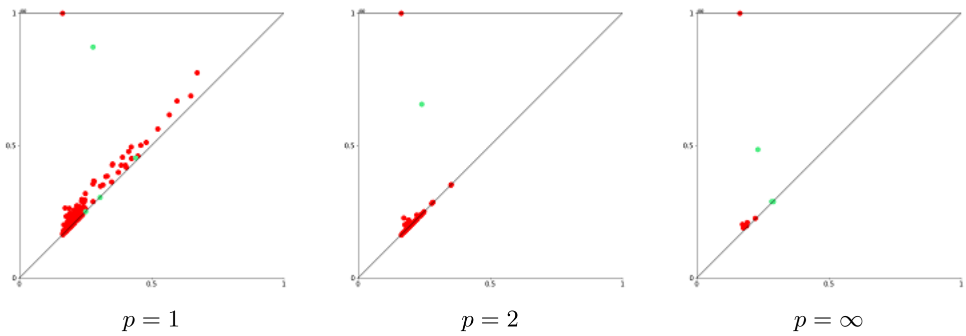
$$d_i(\mathbb{W}[X], \mathbb{W}[Y]) \leq m^{-\frac{1}{2}} W_2(X, \Gamma) + m^{-\frac{1}{2}} W_2(\Gamma, \Omega) + m^{-\frac{1}{2}} W_2(\Omega, Y) + c(\Gamma, m, p) + c(\Omega, m, p).$$

Moreover, with $\Omega = Y$, we obtain

$$d_i(\mathbb{W}[X], \mathbb{W}[\Gamma]) \leq m^{-\frac{1}{2}} W_2(X, \Gamma) + m^{-\frac{1}{2}} W_2(\Gamma, \Omega) + c(\Gamma, m, p) + c(\Omega, m, p).$$

Notice that when $p = 1$, the constant $c(\Gamma, m, p)$ is equal to the constant $c(\Gamma, m)$ defined in Subsection 4.3, and we recover Theorem 15 in homology.

As an illustration of these results, we represent in Figure 7 the persistence diagrams associated to the filtration $\text{Rips}(\mathcal{W}[X])$ for several values of p. The point cloud X is the one defined in the example page 10. Observe that, as stated in Proposition 8, the number of red points (homology in dimension 0) is non-increasing with respect to p.



■ **Figure 7** Persistence diagrams of the simplicial filtrations $\text{Rips}(\mathcal{W}[X])$ for several values of p.

5 Conclusion

In this paper we have introduced the DTM-filtrations that depend on a parameter $p \geq 1$. This new family of filtrations extends the filtration introduced in [3] that corresponds to the case $p = 2$.

The established stability properties are, as far as we know, of a new type: the closeness of two DTM-filtrations associated to two data sets relies on the existence of a well-sampled underlying object that approximates both data sets in the Wasserstein metric. This makes the DTM filtrations robust to outliers. Even though large values of p lead to persistence diagrams with less points in the 0th homology, the choice of $p = 1$ gives the strongest stability results. When $p > 1$, the interleaving bound is less significant since it involves the diameter of the underlying object, but the obtained bound is consistent with the case $p = 1$ as it converges to the bound for $p = 1$ as p goes to 1.

It is interesting to notice that the proofs rely on only a few properties of the DTM. As a consequence, the results should extend to other weight functions, such that the DTM with an exponent parameter different from 2, or kernel density estimators. Some variants concerning the radius functions in the weighted Čech filtration, are also worth considering. The analysis shows that one should choose radius functions whose asymptotic behaviour look like the one of the case $p = 1$. In the same spirit as in [12, 3] where sparse-weighted Rips filtrations were considered, it would also be interesting to consider sparse versions of the DTM-filtrations and to study their stability properties.

Last, the obtained stability results, depending on the choice of underlying sets, open the way to the statistical analysis of the persistence diagrams of the DTM-filtrations, a problem that will be addressed in a further work.

References

1 Greg Bell, Austin Lawson, Joshua Martin, James Rudzinski, and Clifford Smyth. Weighted Persistent Homology. *arXiv preprint*, 2017. `arXiv:1709.00097`.

2 Mickaël Buchet. *Topological inference from measures*. PhD thesis, Paris 11, 2014.

3 Mickaël Buchet, Frédéric Chazal, Steve Y Oudot, and Donald R Sheehy. Efficient and robust persistent homology for measures. *Computational Geometry*, 58:70–96, 2016.

4 F. Chazal, D. Cohen-Steiner, and Q. Mérigot. Geometric Inference for Probability Measures. *Journal on Found. of Comp. Mathematics*, 11(6):733–751, 2011.

5 Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics, 2016.

6 Frédéric Chazal, Vin De Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173(1):193–214, 2014.

7 Frédéric Chazal and Steve Yann Oudot. Towards persistence-based reconstruction in euclidean spaces. In *Proceedings of the twenty-fourth annual symposium on Computational geometry*, SCG '08, pages 232–241, New York, NY, USA, 2008. ACM.

8 Leonidas Guibas, Dmitriy Morozov, and Quentin Mérigot. Witnessed k-distance. *Discrete & Computational Geometry*, 49(1):22–45, 2013.

9 Fujitsu Laboratories. Estimating the Degradation State of Old Bridges-Fijutsu Supports Ever-Increasing Bridge Inspection Tasks with AI Technology. *Fujitsu Journal*, March 2018. URL: `https://journal.jp.fujitsu.com/en/2018/03/01/01/`.

10 J. Phillips, B. Wang, and Y Zheng. Geometric Inference on Kernel Density Estimates. In *Proc. 31st Annu. Sympos. Comput. Geom (SoCG 2015)*, pages 857–871, 2015.

11 Lee M Seversky, Shelby Davis, and Matthew Berger. On time-series topological data analysis: New data and opportunities. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 59–67, 2016.

12 Donald R. Sheehy. Linear-Size Approximations to the Vietoris-Rips Filtration. *Discrete & Computational Geometry*, 49(4):778–796, 2013.

13 Yuhei Umeda. Time Series Classification via Topological Data Analysis. *Transactions of the Japanese Society for Artificial Intelligence*, 32(3):D–G72_1, 2017.

14 Gudhi: Geometry understanding in higher dimensions. `http://gudhi.gforge.inria.fr/`.